

Predição de Séries Temporais Baseada em Redes Neurais Artificiais

Aluno: João Pedro de Oliveira Pagnan [FEEC/UNICAMP]

Orientador: Prof. Dr. Levy Boccato [FEEC/UNICAMP]

Coorientador: Prof. Dr. Romis Ribeiro de Faissol Attux [FEEC/UNICAMP]

Departamento de Engenharia de Computação e Automação Industrial – DCA

25 de outubro de 2021

Seções

1. Introdução

Previsão de séries temporais de sistemas caóticos

Previsão de séries temporais com redes neurais artificiais

2. Cenários escolhidos

Mapa de Hénon

Mapa logístico

Sistema de Lorenz

Equações de Mackey-Glass

3. Redes neurais utilizadas para a previsão

MLP

LSTM e GRU

ESN

4. Metodologia

Configurações das redes neurais

Análise do melhor valor para K

5. Resultados

6. Análise e Conclusão

Introdução

Previsão de séries temporais de sistemas caóticos

A predição de séries temporais é uma das aplicações mais interessantes do tratamento de informação. O desafio de antecipar padrões de comportamento e construir modelos que sejam apropriados para explicar determinados fenômenos da natureza tem importância para a:

- Biologia,

Introdução

Previsão de séries temporais de sistemas caóticos

A predição de séries temporais é uma das aplicações mais interessantes do tratamento de informação. O desafio de antecipar padrões de comportamento e construir modelos que sejam apropriados para explicar determinados fenômenos da natureza tem importância para a:

- Biologia,
- Economia,

Introdução

Previsão de séries temporais de sistemas caóticos

A predição de séries temporais é uma das aplicações mais interessantes do tratamento de informação. O desafio de antecipar padrões de comportamento e construir modelos que sejam apropriados para explicar determinados fenômenos da natureza tem importância para a:

- Biologia,
- Economia,
- Automação Industrial,

Introdução

Previsão de séries temporais de sistemas caóticos

A predição de séries temporais é uma das aplicações mais interessantes do tratamento de informação. O desafio de antecipar padrões de comportamento e construir modelos que sejam apropriados para explicar determinados fenômenos da natureza tem importância para a:

- Biologia,
- Economia,
- Automação Industrial,
- Meteorologia,

Introdução

Previsão de séries temporais de sistemas caóticos

A predição de séries temporais é uma das aplicações mais interessantes do tratamento de informação. O desafio de antecipar padrões de comportamento e construir modelos que sejam apropriados para explicar determinados fenômenos da natureza tem importância para a:

- Biologia,
- Economia,
- Automação Industrial,
- Meteorologia,
- e várias outras áreas da ciência.

Introdução

Previsão de séries temporais de sistemas caóticos

Uma classe de sistemas dinâmicos particularmente relevante dentro do contexto de modelagem e predição de séries temporais está ligada à ideia de dinâmica caótica. Apesar de serem determinísticos, esses sistemas são extremamente sensíveis às condições iniciais^[1].

Introdução

Previsão de séries temporais de sistemas caóticos

Tal característica intensifica o desafio de predizer as séries temporais originadas por eles, pois uma pequena incerteza na medida afetará toda a previsão.

Introdução

Previsão de séries temporais com redes neurais artificiais

Tendo em vista o desempenho de modelos não-lineares para previsão de diversas séries temporais^[2], optamos por estudar a aplicabilidade de redes neurais artificiais à previsão de séries relacionadas a sistemas com dinâmica caótica.

Esta pesquisa comparou o desempenho de quatro arquiteturas de redes neurais artificiais:

- *Multilayer Perceptron (MLP)*^[3],

Introdução

Previsão de séries temporais com redes neurais artificiais

Tendo em vista o desempenho de modelos não-lineares para previsão de diversas séries temporais^[2], optamos por estudar a aplicabilidade de redes neurais artificiais à previsão de séries relacionadas a sistemas com dinâmica caótica.

Esta pesquisa comparou o desempenho de quatro arquiteturas de redes neurais artificiais:

- *Multilayer Perceptron (MLP)*^[3],
- *Long Short-Term Memory (LSTM)*^[2],

Introdução

Previsão de séries temporais com redes neurais artificiais

Tendo em vista o desempenho de modelos não-lineares para previsão de diversas séries temporais^[2], optamos por estudar a aplicabilidade de redes neurais artificiais à previsão de séries relacionadas a sistemas com dinâmica caótica.

Esta pesquisa comparou o desempenho de quatro arquiteturas de redes neurais artificiais:

- *Multilayer Perceptron (MLP)*^[3],
- *Long Short-Term Memory (LSTM)*^[2],
- *Gated Recurrent Unit (GRU)*^[4],

Introdução

Previsão de séries temporais com redes neurais artificiais

Tendo em vista o desempenho de modelos não-lineares para previsão de diversas séries temporais^[2], optamos por estudar a aplicabilidade de redes neurais artificiais à previsão de séries relacionadas a sistemas com dinâmica caótica.

Esta pesquisa comparou o desempenho de quatro arquiteturas de redes neurais artificiais:

- *Multilayer Perceptron (MLP)*^[3],
- *Long Short-Term Memory (LSTM)*^[2],
- *Gated Recurrent Unit (GRU)*^[4],
- *Echo State Network (ESN)*^[5].

Cenários escolhidos

A comparação foi realizada considerando quatro cenários, sendo dois destes a tempo discreto e dois a tempo contínuo, sendo que foi gerado um conjunto de 5000 amostras para cada sistema analisado. Os cenários escolhidos foram as séries temporais dos sistemas caóticos:

- Mapa de Hénon^[6],

Cenários escolhidos

A comparação foi realizada considerando quatro cenários, sendo dois destes a tempo discreto e dois a tempo contínuo, sendo que foi gerado um conjunto de 5000 amostras para cada sistema analisado. Os cenários escolhidos foram as séries temporais dos sistemas caóticos:

- Mapa de Hénon^[6],
- Mapa logístico^[7],

Cenários escolhidos

A comparação foi realizada considerando quatro cenários, sendo dois destes a tempo discreto e dois a tempo contínuo, sendo que foi gerado um conjunto de 5000 amostras para cada sistema analisado. Os cenários escolhidos foram as séries temporais dos sistemas caóticos:

- Mapa de Hénon^[6],
- Mapa logístico^[7],
- Sistema de Lorenz^[8],

Cenários escolhidos

A comparação foi realizada considerando quatro cenários, sendo dois destes a tempo discreto e dois a tempo contínuo, sendo que foi gerado um conjunto de 5000 amostras para cada sistema analisado. Os cenários escolhidos foram as séries temporais dos sistemas caóticos:

- Mapa de Hénon^[6],
- Mapa logístico^[7],
- Sistema de Lorenz^[8],
- Equações de Mackey-Glass^[9].

Cenários escolhidos

Mapa de Hénon

Esse sistema foi proposto por Michel Hénon em 1976 como um modelo simplificado de uma seção de Poincaré do atrator de Lorenz, sendo descrito pelas equações abaixo^[6]:

$$x[n + 1] = y[n] + 1 - a \cdot (x[n])^2 \quad (1a)$$

$$y[n + 1] = b \cdot x[n] \quad (1b)$$

Cenários escolhidos

Mapa de Hénon

Para esta pesquisa, foram utilizados os valores usuais para os parâmetros a e b . Logo, têm-se que $a = 1.4$ e $b = 0.3$. Além disso, neste e nos outros sistemas. A figura 1 mostra a série temporal referente à variável x e o atrator obtido com a simulação para $[x[0] \ y[0]]^T = [1 \ 0]^T$.

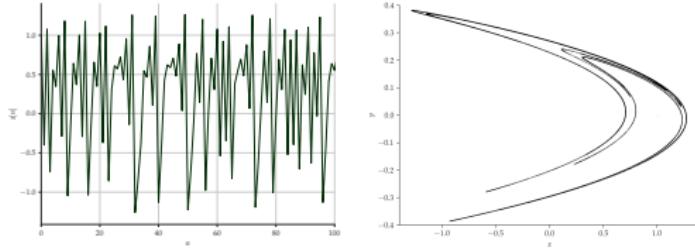


Figura: À esquerda, as cem primeiras iterações da série temporal em x do mapa de Hénon e, à direita, o atrator correspondente à simulação

Cenários escolhidos

Mapa logístico

Descrito em 1976 por Robert May, o mapa logístico representa uma das formas de modelar a população de uma determinada espécie em certos instantes de tempo^[7]. A equação a diferenças que descreve este sistema pode ser vista abaixo:

$$x[n + 1] = r \cdot x[n] \cdot (1 - x[n]) \quad (2)$$

Cenários escolhidos

Mapa logístico

Como o estudo visa analisar o desempenho para sistemas caóticos, foi utilizado $r = 3.86$, que, conforme será visto no diagrama de bifurcação abaixo, faz com que a série temporal dada pela equação (2) opere em caos.

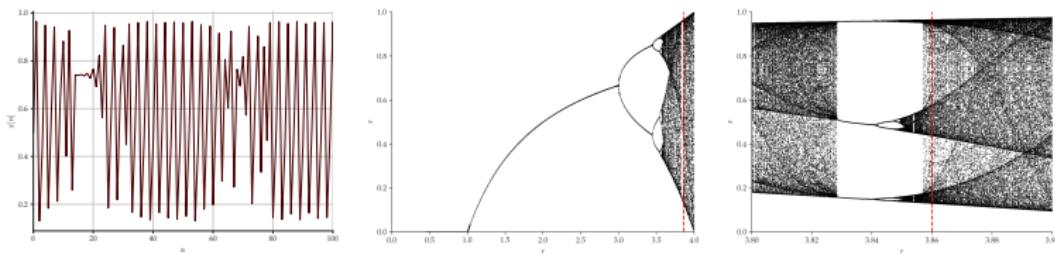


Figura: À esquerda, as cem primeiras iterações da série temporal do mapa logístico e, à direita, o diagrama de bifurcação deste sistema

Cenários escolhidos

Sistema de Lorenz

Num dos primeiros grandes trabalhos envolvendo a noção de regime caótico, sendo considerado por muitos a pesquisa que inaugurou a área^[10], Lorenz modela o fluxo de um fluido em um volume uniformemente aquecido na camada inferior e uniformemente resfriado na camada superior^[8]:

$$\frac{dx}{dt} = -\sigma \cdot (x - y) \quad (3a)$$

$$\frac{dy}{dt} = x \cdot (\rho - z) - y \quad (3b)$$

$$\frac{dz}{dt} = x \cdot y - \beta \cdot z \quad (3c)$$

Cenários escolhidos

Sistema de Lorenz

Para a simulação numérica foi considerado que $\sigma = 10$, $\beta = \frac{8}{3}$, $\rho = 28$ e foi utilizado $dt = 0.01$. A figura 3 mostra a série temporal na variável x e o atrator de Lorenz para a condição inicial $[x(0) \ y(0) \ z(0)]^T = [0.1 \ 0 \ 0]^T$.

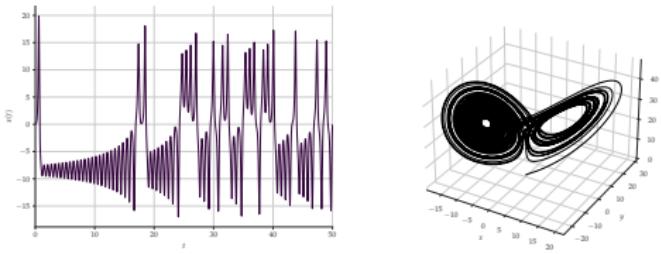


Figura: À esquerda, a série temporal em x do sistema de Lorenz simulado e, à direita, o diagrama de fases correspondente à simulação

Cenários escolhidos

Equações de Mackey-Glass

O último sistema caótico simulado está associado às equações de Mackey-Glass, as quais modelam o controle hormonal da produção de células brancas do sangue e podem ser vistas abaixo^[9]:

$$\frac{dP(t)}{dt} = \frac{\beta_0 \cdot \theta^n}{\theta^n + P(t - \tau)^n} - \gamma \cdot P(t) \quad (4a)$$

$$\frac{dP(t)}{dt} = \frac{\beta_0 \cdot \theta^n \cdot P(t - \tau)}{\theta^n + P(t - \tau)^n} - \gamma \cdot P(t) \quad (4b)$$

Cenários escolhidos

Equações de Mackey-Glass

A equação (4b) exibe comportamento caótico para valores mais altos de τ . Para a simulação numérica, foi utilizado $n = 10$, $\gamma = 0.1$, $\beta = 0.2$, $\theta = 1$, $\tau = 22$, $dt = 1.0$ e $P(0^-) = 0.1$, gerando novamente 5000 amostras. A série e o atrator obtidos podem ser vistos na figura 4.

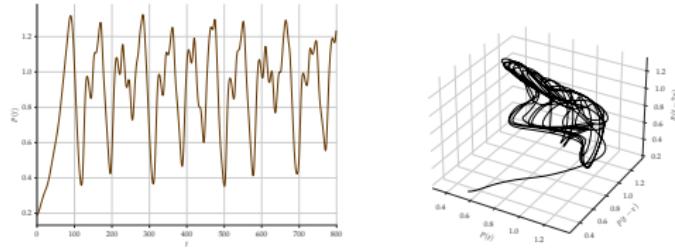


Figura: À esquerda, a série temporal da equação (4b) exibida de $t = 0$ a $t = 800$ e, à direita, o atrator correspondente à simulação

Redes neurais utilizadas para a previsão

MLP

As redes MLP, como o nome já diz, são compostas por múltiplas camadas de neurônios artificiais do tipo *perceptron*^[3]. Cada neurônio recebe os atributos $x_i, i = 1, \dots, m$ de entrada e os pondera por pesos w_i . Em seguida, é aplicada uma função de ativação $\varphi(\cdot)$ sobre esta soma ponderada, acrescida de um termo de *bias*, conforme indica a equação (5).

$$y = \varphi \left(\sum_{i=1}^m w_i x_i + w_0 \right) = \varphi \left(\sum_{i=0}^m w_i x_i \right) = \varphi (\mathbf{w}^T \cdot \mathbf{x}) \quad (5)$$

Redes neurais utilizadas para a previsão MLP

Uma rede neural MLP é composta por um número arbitrário N_L de camadas com n neurônios do tipo *perceptron*, com a característica de que as saídas dos neurônios da l -ésima camada são propagadas para a frente, servindo como as entradas de todos os neurônios da camada seguinte ($l + 1$). A figura 5 apresenta a estrutura típica das redes MLP.

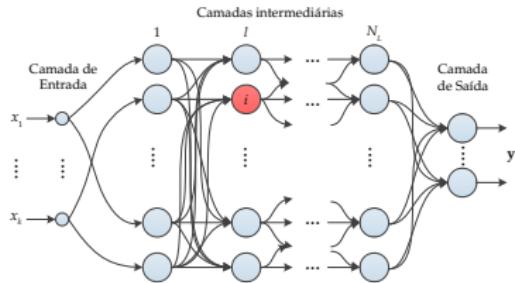


Figura: Estrutura típica de uma rede MLP (figura adaptada de^[11])

Redes neurais utilizadas para a previsão

MLP

O processo de treinamento de uma rede neural artificial normalmente é realizado com sequências de vetores de entrada x , chamadas de *mini-batch*, sendo que um período de treinamento é chamado de época^[12].

Os pesos sinápticos w são ajustados com um processo iterativo de forma a minimizar uma função custo $J(w)$ que representa uma medida do erro entre as saídas geradas pela rede e as saídas desejadas.

Redes neurais utilizadas para a previsão

LSTM e GRU

As redes recorrentes têm estruturas computacionais que podem armazenar os estados anteriores dos neurônios, possuindo também portas não-lineares que regulam o fluxo de informação de entrada e de saída da célula computacional^[13].

A diferença entre uma célula LSTM e uma célula GRU é a presença de um vetor de longo prazo c e um vetor de curto prazo h na LSTM, conforme indicado na figura 6.

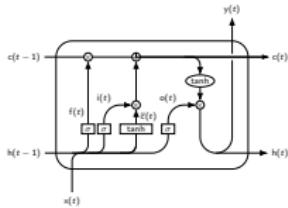


Figura: Estrutura de uma célula LSTM.

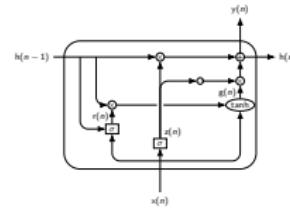


Figura: Estrutura de uma célula GRU.

Redes neurais utilizadas para a previsão

LSTM e GRU

À semelhança das redes MLP, o treinamento das redes LSTM e GRU também é realizado através de algoritmos de otimização baseados em derivadas da função custo; a diferença é que agora é necessário propagar as derivadas ao longo da estrutura e, também, ao longo do tempo devido às realimentações^[12].

Redes neurais utilizadas para a previsão ESN

As redes neurais com estados de eco também são modelos recorrentes para processamento da informação, à semelhança da LSTM e da GRU. No entanto, apresentam um modo de operação e um esquema de treinamento bem diferentes^[11].

A figura 8 apresenta a estrutura interna de uma rede ESN.

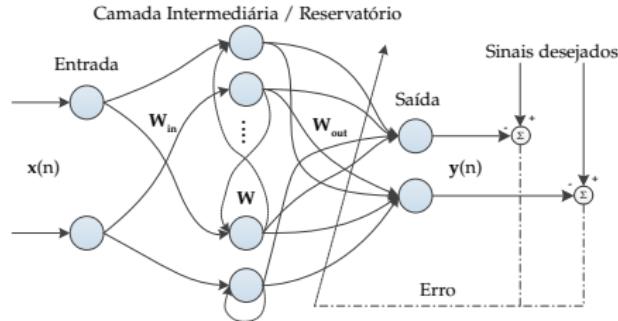


Figura: Estrutura típica de uma rede ESN (figura adaptada de^[11]).

Redes neurais utilizadas para a previsão ESN

O grande diferencial prático desta rede é que os pesos das conexões recorrentes dentro do reservatório W são ajustados com valores fixos antes do treinamento da camada de saída. Esses parâmetros são determinados tendo em vista a propriedade de estados de eco, existindo alguns métodos simples para a criação aleatória dos pesos e que asseguram a propriedade, conforme demonstrado em [5].

Metodologia

Iniciamos a análise através de um processo de busca em grade para determinarmos os parâmetros ótimos das redes neurais em cada cenário.

Em seguida, avaliamos a progressão do erro quadrático médio (EQM) em função do número de amostras de entrada do modelo preditor (chamado de K).

Metodologia

Por fim, comparamos qual foi a média e o desvio padrão do EQM com o melhor valor de K de cada modelo nos quatro cenários. O horizonte de predição utilizado foi $L = 3$. Assim, iremos predizer o valor da terceira iteração à frente do valor atual da série temporal.

Metodologia

Configurações das redes neurais

Para cada arquitetura, um conjunto de valores candidatos foi gerado para cada hiperparâmetro e todas as combinações possíveis foram avaliadas tendo em vista o desempenho do modelo em dados de validação, utilizando um processo de busca em grade.

Mais detalhes podem ser vistos no artigo enviado ao congresso.

Metodologia

Análise do melhor valor para K

Com as melhores configurações para cada rede e cenário obtidas, foi analisada a progressão do erro quadrático médio para cada valor de K . A faixa de valores para K a ser testada foi determinada utilizando a autocorrelação parcial das séries temporais de cada sistema analisado, levando em consideração os valores de K que têm as autocorrelações parciais mais relevantes.

Metodologia

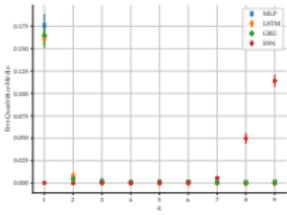
Análise do melhor valor para K

Para realizar esse procedimento, cada rede (com as configurações ótimas) foi treinada utilizando 85% dos dados gerados, sendo que 10% dos dados de treinamento foi utilizado como o conjunto de validação (nas redes MLP, LSTM e GRU). Em seguida, foi avaliado o EQM no conjunto de teste (que corresponde às últimas 750 amostras). Esse processo foi realizado 5 vezes para cada K , obtendo assim um valor médio e desvio-padrão para cada modelo e em cada cenário.

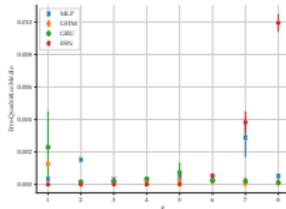
Metodologia

Análise do melhor valor para K

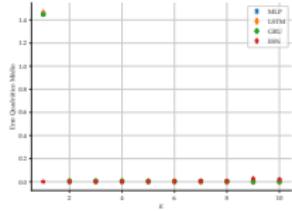
A figura 9 mostra a progressão do EQM observada em cada cenário.



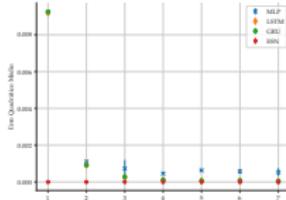
(a) Mapa de Hénon



(b) Mapa logístico



(c) Sistema de Lorenz



(d) Equações de Mackey-Glass

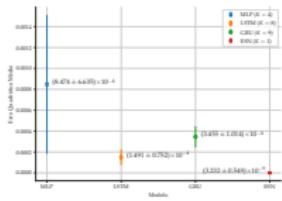
Figura: Progressão do erro quadrático médio para cada valor de K .

Resultados

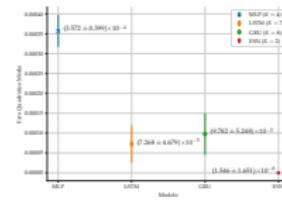
Após identificarmos o melhor valor de K para cada modelo, realizou-se novamente o processo mencionado na seção anterior e foi obtido o EQM para as melhores configurações (parâmetros e K) para cada modelo, em todos os cenários.

Resultados

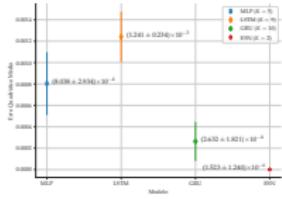
A figura 10 exibe o comparativo dos desempenhos obtidos por cada modelo com as melhores configurações nos cenários testados.



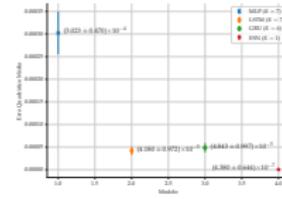
(a) Mapa de Hénon



(b) Mapa logístico



(c) Sistema de Lorenz

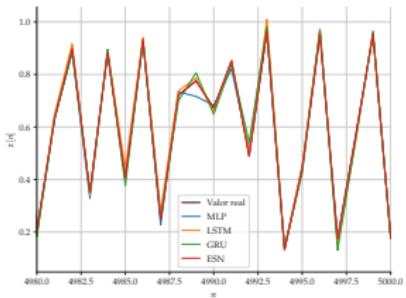


(d) Equações de Mackey-Glass

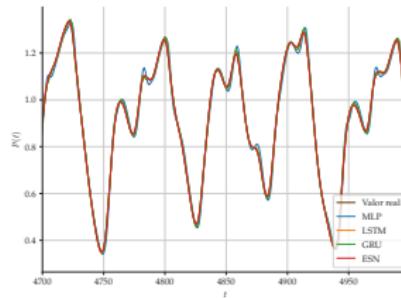
Figura: Comparação do desempenho obtido por cada modelo.

Resultados

A figura 11 mostra uma comparação da predição em si de cada modelo nos cenários onde a diferença foi mais perceptível visualmente.



(a) Mapa logístico



(b) Equações de Mackey-Glass

Figura: Comparação da predição realizada por cada modelo nos cenários do mapa logístico e das equações de Mackey-Glass.

Análise e Conclusão

Analisando os resultados obtidos, tiramos algumas conclusões importantes:

- A rede MLP foi consideravelmente pior do que as redes recorrentes,

Análise e Conclusão

Analizando os resultados obtidos, tiramos algumas conclusões importantes:

- A rede MLP foi consideravelmente pior do que as redes recorrentes,
- Dentre as redes recorrentes, a ESN obteve um EQM bem inferior ao obtido pela LSTM e pela GRU,

Análise e Conclusão

Analizando os resultados obtidos, tiramos algumas conclusões importantes:

- A rede MLP foi consideravelmente pior do que as redes recorrentes,
- Dentre as redes recorrentes, a ESN obteve um EQM bem inferior ao obtido pela LSTM e pela GRU,
- **A ESN atingiu patamares de erro cerca de 100 ou até 10000 vezes menores que os outros modelos.**

Análise e Conclusão

O desempenho inferior da rede MLP com relação às redes recorrentes provavelmente decorre do fato de que a relação temporal presente na LSTM, GRU e ESN auxilia na modelagem da dinâmica da série temporal. Já o pior desempenho da LSTM na série temporal do sistema de Lorenz provavelmente está relacionado aos efeitos estocásticos presentes no ajuste dos pesos sinápticos dessa rede neural que, conforme indicado em^[14], é uma dificuldade em seu treinamento.

Análise e Conclusão

Considerando que a ESN obteve, de longe, o melhor desempenho, e que o treinamento desta arquitetura é bem menos custoso computacionalmente se comparado às demais, pode-se concluir que esta rede é uma boa alternativa para futuros estudos de modelos preditores de séries temporais.

Análise e Conclusão

Além disso, conforme mostrado em outros trabalhos como^[5,11,15], a ESN também é uma boa ferramenta para outras tarefas de extração de informação, como em equalização de canais e separação de fontes.

Análise e Conclusão

Como sugestão de trabalhos futuros, pode ser avaliada a eficácia da ESN em reconstruir atratores através de séries temporais experimentais de sistemas caóticos. Se o desempenho para essa tarefa for tão bom quanto o obtido na predição das séries estudadas, a ESN pode tornar-se uma ferramenta ainda mais poderosa para a modelagem de sistemas com dinâmica caótica.

Referências I

- [1] N. Fiedler-Ferrara and C. P. C. do Prado, *Caos: uma introdução*. Editora Blucher, 1994.
- [2] J. T. Connor, R. D. Martin, and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 240–254, 1994.
- [3] F. Rosenblatt, "The perceptron: a probabilistic model for information storage and organization in the brain.," *Psychological review*, vol. 65, no. 6, p. 386, 1958.
- [4] K. Cho, B. Van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
- [5] H. Jaeger, "Echo state network," *scholarpedia*, vol. 2, no. 9, p. 2330, 2007.
- [6] M. Hénon, "A two-dimensional mapping with a strange attractor," *Communications in Mathematical Physics*, vol. 50, pp. 69–77, feb 1976.
- [7] R. M. May, "Simple mathematical models with very complicated dynamics," *Nature*, vol. 261, pp. 459–467, jun 1976.
- [8] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of atmospheric sciences*, vol. 20, no. 2, pp. 130–141, 1963.
- [9] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.
- [10] J. Gleick, *Chaos: The amazing science of the unpredictable*. Vintage Publishing, 1998.
- [11] L. Boccato, *Novas propostas e aplicações de redes neurais com estados de eco*. Tese (doutorado), Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação, Campinas, SP, 2013.

Referências II

- [12] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.
- [13] S. Haykin, *Neural networks and learning machines*, 3/E. Pearson Education India, 2010.
- [14] K. Doya *et al.*, "Bifurcations in the learning of recurrent neural networks 3," *learning (RTRL)*, vol. 3, p. 17, 1992.
- [15] H. Jaeger and H. Haas, "Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication," *science*, vol. 304, no. 5667, pp. 78–80, 2004.

Muito Obrigado!